REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE dd-mm-yyyy	3. REPORT TYPE AND DATES COVERED
	07-21-2003	Final 01-08-1999 to 31-12-2002

4. TITLE AND SUBTITLE
Final Report on Nonparametric Function Estimation and Visualization Application to C2

5. FUNDING NUMBERS
DAAD19-99-1-0314

6. AUTHOR(S)
Edward J. Wegman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
George Mason University, MS 4A7
4400 University Drive
Fairfax, VA 22030-4444

8. PERFORMING ORGANIZATION REPORT NUMBER Final 2003-01

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
~~39574~~
40276.1-MA

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This project focused on the development of fast, accurate density estimation procedures. The methods raised basic research issues as to the implementation, computational complexity, visualization, and optimization of estimators in this class. In addition to being useful in a direct role, it is argued that density estimation plays a crucial role in clustering algorithms, discriminant methods and pattern recognition. All of these methods are used extensively in Situation and Informational Awareness and Understanding and in Monitoring and Discovery Processes. In addition, because of their intuitive appeal and ease in understanding, visually rendered density and function estimators provide a natural format for human-computer interactions with decision makers. This report describes results related implementation, computational complexity, visualization, optimization and application of recursive orthonormal density estimators.

14. SUBJECT TERMS
visual data mining
quantization
clustering
nonparametric density estimation

15. NUMBER OF PAGES
7

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

SF 298 Continuation Sheet

FINAL REPORT

ARO PN 40276-MA

CONTRACT NO. DAAD19-99-1-0314

Period of Performance: 1 August 1999 to 31 December 2002

Title: Nonparametric Function Estimation and Visualization with Applications to C2

PI: Edward J. Wegman, George Mason University

Papers Published Under ARO Sponsorship

Special Issues or Books

Wegman, E. (ed.) (2000) *On the Eve of the 21st Century*, Special Issue of *Computational Statistics and Data Analysis*, guest edited, Vol. 32, Nos. 3 and 4.

Wegman, E. and Martinez, Y. (2000) *Computing Science and Statistics: Proceedings of the 32nd Symposium on the Interface* (issued as a CD), Fairfax Station, VA: Interface Foundation of North America, Inc.

National Research Council, *Naval Forces' Capability for Theater Missile Defense*, (Wegman was a member of NAS-NRC authoring committee), Washington, DC: National Academy Press, 2001

Manuscripts Submitted but not yet Published

Wegman, E. J. and Dorfman, A. H. "Visualizing cereal world," to appear *Computational Statistics and Data Analysis*

Chow, Winston, Wegman, E. J. "Modeling continuous time series driven by fractional Gaussian noise," to appear *Institute for Mathematics and its Applications Monographs*

Wegman, E. J. "On some statistical methods for parallel computation," to appear *Handbook of Parallel Computing and Statistics*

Solka, J. L., Wegman, E. J., and Marchette, D. J. "Data mining strategies for detection of chemical warfare agents," to appear *Statistical Data Mining and Knowledge Discovery*

Marchette, D. J., Wegman, E. J., and Priebe, C.E. "A fast algorithm for approximating the dominating set of a class cover digraph," submitted to *Journal of Computational and Graphical Statistics*

Moustafa, R. E. A. and Wegman, E. J. "On some generalizations of parallel coordinates," submitted to *Computational Statistics*

Dorfman, A. H., Lent, Janice, Leaver, S. G. and Wegman, E. J. "On sample survey designs for consumer price indexes," with submitted.

Khumbah, N.-A. and Wegman, E. J. "Data compression by geometric quantization," to appear *Recent Advances and Trends in Nonparametric Statistics.*

Marchette, D. J. and Wegman, E. J. "Statistical analysis of network data for cybersecurity," to appear *Chance.*

Papers Published in Peer Review Journals

Wilhelm, A. F. X, Wegman, E. J. and Symanzik, J. (1999) "Visual clustering and classification: The Oronsay particle size data set revisited," *Computational Statistics*, 14(1), 109-146.

Wegman, E. J. (1999) "Visions: The evolution of statistics," *Research in Official Statistics*, 2(1), 7-19.

Chen, J. X., Fu, X., and Wegman, E. J. (1999) "Real-time simulation of dust behaviors generated by a fast traveling vehicle," *ACM Transactions on Modeling and Computer Simulation*, 9(2), 81-104.

Wegman, E. J. (2000) "On the eve of the 21st century: Statistical science at a crossroads," *Computational Statistics and Data Analysis*, 32, 239-243.

Wegman, E. J. (2000) "Visions: New techniques and technologies in statistics," *Computational Statistics*, 15, 133-144.

Martinez, W. and Wegman, E. (2000) "An alternative criterion useful for finding E-optimal designs," *Statistics and Probability Letters*, 47, 325-328.

Wegman, E. J. (2000) Book Review of *The Grammar of Graphics* by Leland Wilkinson, *Journal of the American Statistical Association*, 95(451), 1009-1010.

Wegman, E. J. (2000) "Affordable environments for 3D collaborative data visualization," *Computation in Science and Engineering*, 2(6), 68-72, 74.

Chen, J. X., Wang, J. and Wegman, E. J. (2000) "Physical model of dust behaviors behind a moving object," *International Journal of Applied Science and Computations*, 7(2), 1-12.

Wegman, E. J. and Luo, Q. (2002) "On methods of computer graphics for visualizing densities," *Journal of Computational and Graphical Statistics*, 11(1), 137-162.

Wegman, E. J. and Symanzik, J. (2002) "Immersive projection technology for visual data mining," *Journal of Computational and Graphical Statistics*, 11(1), 163-188.

Wegman, E. J. and Solka, J. L. (2002) "On some mathematics for visualizing high dimensional data," *Sanhkya (A)*, 64(2), 429-452.

Wegman, E. J. (2003) "Visual data mining," *Statistics in Medicine*, 22, 1383-1397.

Papers Published in Non-Peer Reviewed Journals or Conference Proceedings

Wegman, E. J., J. Symanzik, J.P. Vandersluis, Q. Luo, F. Camelli, A. Dzubay, X. Fu, N-A. Khumbah, R. Moustafa, R. Wall and Y. Zhu (1999) "The MiniCAVE - A voice-controlled IPT environment," *Proceedings of the Third International Immersive Projection Technology Workshop*, (H.-J. Bullinger and O. Riedel, eds.), Springer-Verlag, Berlin, 179-190.

Wegman, E. J. (1999) "Data mining and visualization: some strategies," *Bulletin of the International Statistical Institute*, Tome LVIII, Book 3, 223-226.

Moustafa, R. E. A. and Wegman, E. J. (1999) "Using genetic algorithms (GAs) for the gene mapping problem," *Computing Science and Statistics*, 31, 487-492.

Wegman, E. J. and Solka, J. L. (1999) "Implications of distance learning methodologies for statistical education," *ASA Proceedings of the Sections on Statistical Education, Teaching Statistics in the Health Sciences, and Statistical Consulting*, 13-16.

Moustafa, R. E. A., DeJong, K. and Wegman, E. J. (1999) "Adaptive numerical approximation based on genetic algorithms," *Proceedings of the 1999 Genetic and Evolutionary Computing Conference (GECCO).*

Moustafa, R. and Wegman, E. (2000) "Mining evolutionary models to multidimensional scaling of gene measurements," *Computing Science and Statistics*, 32, /HTMLProceedings/RMoustafa/moustafa.pdf (CD-based publication).

Moustafa, R. and Wegman, E. (2000) "A GA-based method for function approximation using adaptive interpolation," *Proceedings of the 2000 Genetic and Evolutionary Computing Conference (GECCO).*

Wegman, E. (2000) "Authenticating Vulnerability Measurements", *Computing Science and Statistics*, 32, 284-293.

Wegman, E. J. and Symanzik, J. (2001) "Data visualization and exploration via virtual reality: An overview," *Bulletin of the International Statistical Institute*, LIX(2), 76-79.

Dorfman, A., Lent, J., Leaver, S. and Wegman, E. (2001) "On sample survey designs for consumer price indexes," B*ulletin of the International Statistical Institute*, LIX(2), 421-424.

Wegman, E. J., Symanzik, J., Braverman, A. and Luo, Q. (2002) "New applications of the image grand tour," *Computing Science and Statistics*, 34, 500-512.

Martinez, A. and Wegman, E. J. (2002) "A text stream transformation for semantic-based clustering," *Computing Science and Statistics*, 34, 184-203.

Papers Presented at Meetings but not Published in Conference Proceedings

Wegman, E. J. (1999) Distinguished lecture series: Two lectures on data mining and visualization, Utah State University, Logan, UT

Wegman, E. J. (1999) "Roundtable: New Graphics Environments," Joint Statistical Meetings, Baltimore, August, 1999

Wegman, E. J. and Solka, J. L. (1999) "Statistical Education at the Interface: Distance Learning/Computers in the Classroom," Joint Statistical Meetings, Baltimore, August, 1999

Wegman, E. J. (1999) "Data Mining and Visualization," International Statistical Institute, Helsinki, Finland, August, 1999

Wegman, E. J. (1999) "Visual Data Mining," Keynote talk at ACAS 99, The Army Conference on Applied Statistics, West Point, NY, October, 1999

Wegman, E. J. (1999) "Assessing Vulnerability Measurements," INFORMS, Philadelphia, PA, November, 1999

Wegman, E. J. (1999) Two Lectures, Troisieme Cycle de la Suisse Romande, Ecole Polytechnique Federale de Lausanne, Switzerland, December, 1999

Wegman, E. (2000) "Visual Data Mining," Conference in Honor of the 80th Birthday of Professor C. R. Rao, Austin, TX, March, 2000.

Wegman, E. (2000) "Statistical Data Mining," Two-day Short Course Organized by the Washington Statistical Society, Washington, DC, April, 2000.

Wegman, E. (2000) "Data Reduction by Quantization," 5th World Congress of the Bernoulli Society and the Institute of Mathematical Statistics, Guanajuato, Mexico, May, 2000.

Wegman, E. and Luo, Q. (2000) "The MiniCAVE and Crystal Vision DataMining Software," Invited Technical Demonstration, Joint Statistical Meetings, Indianapolis, IN, August, 2000.

Wegman, E. (2000) "CrystalVision: A New Visual DataMining Software," Joint Statistical Meetings, Indianapolis, IN, August, 2000.

Luo, Q. and Wegman, E. (2000) "Visual Data Mining," Joint Statistical Meetings, Indianapolis, IN, August, 2000.

Wegman, E. (2000) "Multivariate Density Estimation: Adaptive Mixtures" and "Multivariate Density Estimation: Geometric Approaches," Workshop on Nonparametric Model Building, Splines and other Smoothing Techniques, State College, PA, October, 2000.

Wegman, E. (2000) "Visual Data Mining," Graduiertenkolleg "Angewandte Statistik," Herbstkolloquium, Dortmund, Germany, November, 2000.

Wegman, E. (2000) "Crystal Vision: A New Visual DataMining Software," Conference on Data Mining and Statistics, Augsburg, Germany, November, 2000.

Wegman, E. J. (2001) "Data Reduction by Quantization," Nonparametrics in Large, Multidimensional Data Mining Conference, Dallas, TX, January, 2001

Wegman, E. J. (2001) "Visual Data Mining," 8th Biennial CDC/ATSDR Statistics Symposium, Atlanta, GA, January, 2001

Wegman, E. J. (2001) Short Course on Statistical Data Mining, ENAR Meeting, Charlotte, NC, March, 2001

Wegman, E. J. (2001) Five Lectures on Geometry, Visualization and Data Mining, University of Aalborg, Denmark, May, 2001

Wegman, E. J. (2001) "Visual Data Mining," Keynote Talk, Danish Society of Theoretical Statistics, Aalborg, Denmark, May, 2001

Wegman, E. J. (2001) "Visualizing Cereal World," DataViz II Workshop, Fairfax, VA, May, 2001

Wegman, E. J. (2001) Short Course on Statistical Data Mining, Interface '01, Orange County, CA, June 2001

Wegman, E. J. (2001) "Data Reduction by Quantization," Joint Statistical Meetings, Atlanta, GA, August, 2001

Wegman, E. J. (2001) "Pixel Tours," IMA Workshop on Geophysics and Statistics, Minneapolis, MN, November, 2001

Wegman, E. J. (2001) "Pixel Tours," American Geophysical Union Meeting, San Francisco, CA, December, 2001

Wegman, E. J. (2002) "Collaborative Visualization Environments," (Best of JCGS Session), Interface 2002, Montreal, Quebec, Canada, April 2002

Wegman, E. J. (2002) Six Lectures on Data Mining, Finnish Summer School in Probability, Lahti, Finland, June 2002

Wegman, E. J. (2002) "Visual Data Mining," Keynote Talk, C. Warren Neel Conference on Statistical Data Mining & Knowledge Discovery, Knoxville, TN, June, 2002

Wegman, E. J. (2002) "Smoothings, Ridges, and Bumps," Joint Statistical Meetings, New York, NY, August, 2002

Wegman, E. J. (2002) "On Some Generalizations of Parallel Coordinate Plots," Seeing a Million – A Data Visualization Workshop, Rain am Lech (nr. Munich), Germany, October, 2002

Wegman, E. J. (2002) "Visual Data Mining," Keynote Talk, M2002 SAS Data Mining Conference, Cary, NC, October 2002

Wegman, E. J. and Solka, J. L. (2002) Short Course on Statistical Data Mining, Army Conference on Applied Statistics, Raleigh, NC, 2002

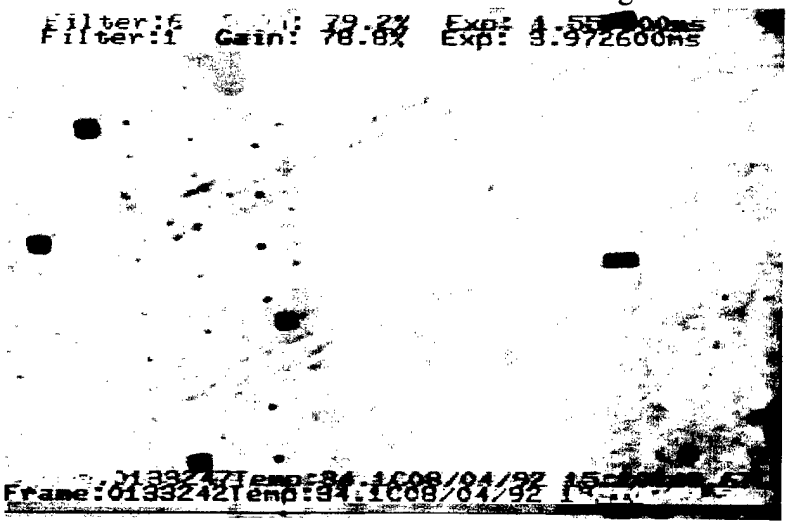
Scientific Personnel Supported and Awards

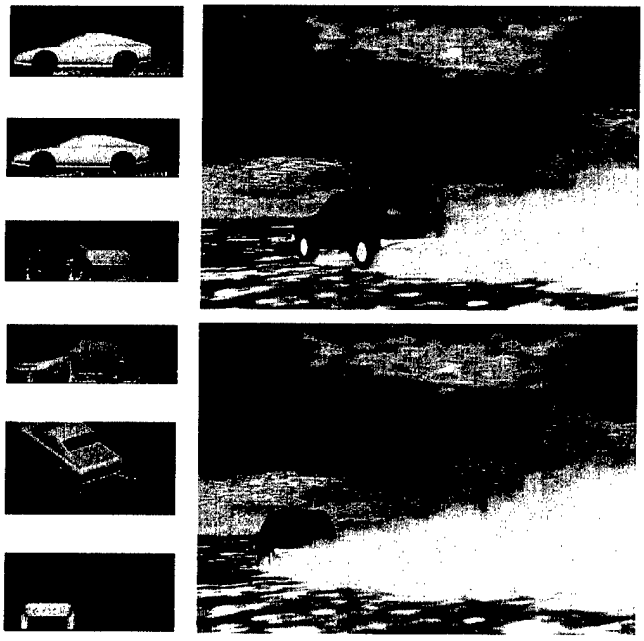
Edward J. Wegman, Awarded Army Wilks Award
Nkem-Amin "Martin" Khumbah, earned Ph.D.
Rida E. A. Moustafa, earned Ph.D.

Dr. Wegman was selected as a NSF/ASA/BLS Senior Faculty Fellow, Spring 2000. He was also elected Chair, Statistical Graphics Section of the American Statistical Association for calendar year 2000.

Scientific Progress and Accomplishments

During the period of performance, progress was made on several fronts. The "Grand Tour" algorithm was extended to hybrid data sets generated by image analysis algorithms. This will extend the possibilities of automatic data analysis and target hunting on a real time basis. It will enlarge the capabilities of the commander in the field when attempting to coordinate and organize incoming data of several types. The image to the right illustrates the ability to use this image grand tour technique to combine images from distinct spectral bands to identify minefields (the three near vertical rows of dots on the left hand side of the image), which are not visible in the individual images.

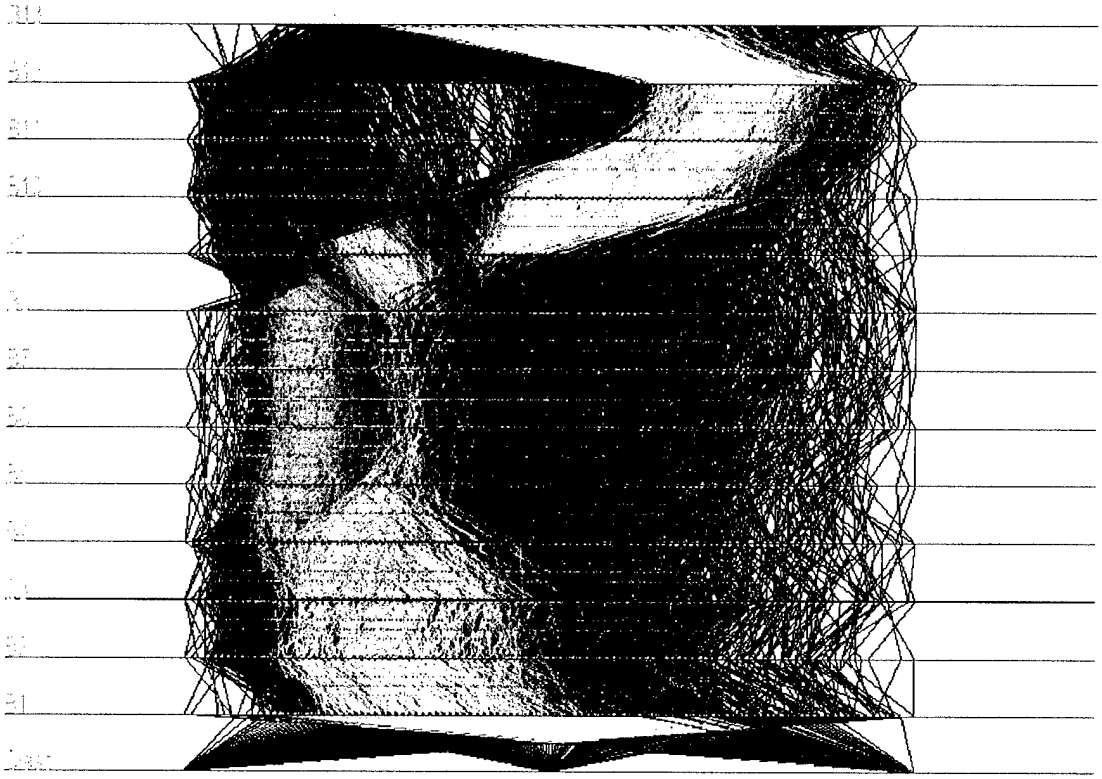




The utilization of D-optimal designs was made possible for robust estimation of multivariate location and scatter. The mathematical basis involves a penalty function approach for simplifying adaptive mixtures and density estimates. This has been completed and programming is being completed at the present time.

Finally, The new algorithms have been tested on several real data sets including the detection and enhancement of the behavior of dust behind moving vehicles in the field, and the effect of dust particles on in a networked virtual environment. See image at left. This is highly appropriate for soldier training in a virtual environment.

The most significant work, which represents the culmination of a long series of research efforts was our paper entitled "Visual data mining." This paper describes a combination of tools developed through ARO sponsorship over an extended period of time and describes techniques for using these tools, including visual techniques for density estimation, rapid data editing, inverse regression, tree-structured decision rules, variable selection and dimension reduction, clustering, classification and discrimination, and outlier and unique event detection.



Parallel Coordinate Display used for Dimension Reduction in the Rapid Detection of Chemical Warfare Agents

Several of the most important papers mentioned above are reproduced on the accompanying CD.